PCT / SE 2004 / 001378

PRV
PATENT- OCH REGISTRERINGSVERKET
Patentavdelningen

REC'D 11 OCT 2004
WIPO PCT

Intyg
Certificate

Härmed intygas att bifogade kopior överensstämmer med de handlingar som ursprungligen ingivits till Patent- och registreringsverket i nedannämnda ansökan.

Ansökan ingavs ursprungligen på engelska.

This is to certify that the annexed is a true copy of the documents as originally filed with the Patent- and Registration Office in connection with the following patent application.

The application was originally filed in English.

(71) Sökande *AstraZeneca AB, Södertälje SE*
Applicant (s)

(21) Patentansökningsnummer *0302559-0*
Patent application number

(86) Ingivningsdatum *2003-09-25*
Date of filing

Stockholm, 2004-09-29

För Patent- och registreringsverket
For the Patent- and Registration Office

Gunilla Larsson

Avgift
Fee

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

PATENT- OCH REGISTRERINGSVERKET
SWEDEN

Postadress/Adress
Box 5055
S-102 42 STOCKHOLM

Telefon/Phone
+46 8 782 25 00
Vx 08-782 25 00

Telex
17978
PATOREG S

Telefax
+46 8 666 02 86
08-666 02 86

FIELD OF THE INVENTION

The invention relates to methods for producing and using protein or peptide fingerprints, and to protein or peptide biomarkers. These methods and biomarkers may be used in the identification, evaluation, study or monitoring of conditions or diseases, for example to aid the discovery, development or use of drugs to treat those conditions or diseases.

BACKGROUND TO THE INVENTION

Various biological markers, known as biomarkers, have been identified and studied through the application of biochemistry and molecular biology to medical and toxicological situations. A biomarker has been described as "a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention". A biomarker is any identifiable and measurable indicator associated with a particular condition or disease where there is a correlation between the presence or level of the biomarker and some aspect of the condition or disease (including the presence of, the level or changing level of, the type of, the stage of, the susceptibility to the condition or disease, or the responsiveness to a drug used for treating the condition or disease). The correlation may be qualitative, quantitative, or both qualitative and quantitative. Typically a biomarker is a compound, compound fragment or group of compounds. Such compounds may be any compounds found in or produced by an organism, including proteins (and peptides), nucleic acids and other compounds.

Biomarkers have a predictive power, and may be used to predict or detect the presence, level, type or stage of particular conditions or diseases (including the presence or level of particular microorganisms or toxins), the susceptibility (including genetic susceptibility) to particular conditions or diseases, or the response to particular treatments (including drug treatments). It is thought that biomarkers will play an increasingly important role in the future of drug discovery and development, by improving the efficiency of research and development programmes. Biomarkers can be used as diagnostic agents, monitors of disease progression, monitors of treatment and predictors of clinical outcome. For

example, various biomarker research projects are attempting to identify markers of specific cancers and of specific cardiovascular and immunological diseases.

Proteomics (including peptidomics) technologies have been developed to analyse proteins (including peptides). These technologies are applied in a high-throughput mode, generating an enormous amount of data that is analysed using computer systems. Proteins from a biological sample are isolated and separated at a high resolution, for example by chromatographic separations. The set of proteins is then characterised using qualitative and quantitative techniques such as mass spectrometry. The result is a protein (or peptide) fingerprint (a constant, reproducible set of proteins or peptides). Selected proteins/peptides or groups of proteins/peptides may be analysed further to generate protein/peptide profiles. Proteomics is now viewed as the large-scale analysis of the function of genes and is becoming a central field in functional genomics.

Separation of proteins is commonly achieved using gel based techniques. 2D-PAGE (polyacrylamide gel electrophoresis) is currently the principal analytical method for studying the cellular expression of proteins. Instrumental platforms allow almost fully automated operations of 2D-gel analysis. The 2D-gel methods have good sensitivity and resolution for a large fraction of expressed proteins, typically those within a mass range of 10-120 kDa. However the methods have significant limitations in the identification of low abundance/low molecular weight proteins, some of which are present at concentrations as low as a few molecules per cell. Problems of sample loss and/or insufficient recovery have confounded the isolation of low abundance/low molecular weight proteins by 2D-PAGE. In addition, the presence of these proteins can be masked by the higher abundance protein spots. Other classes of proteins that are problematic for 2D-PAGE include acidic, basic, hydrophobic and high molecular weight proteins.

Multidimensional HPLC (High Performance Liquid Chromatography) has been used as a good alternative for separating proteins or peptides unsuited to 2D-PAGE. The protein or peptide mixture is passed through a succession of chromatographic stationary phases or

dimensions which gives a higher resolving power. HPLC is also more flexible than the 2D-gel separation methods since the stationary and mobile phases can be selected for their suitability in resolving specific protein or peptide classes of interest and for compatibility with each other and with downstream mass spectrometric methods of detection and identification. On-line configurations of these types of multi-mechanism separation platforms are known.

Mass spectrometry (MS) is also an essential element of the proteomics field. In fact MS is the major tool used to study and characterise purified proteins in this field. The interface link in proteomics and MS, displaying hundreds or thousands of proteins, is made by gel technology where high resolution can be reached on a single gel. Researchers are successfully harnessing the power of MS to supersede the two-dimensional gels that originally gave proteomics its impetus.

The application and development of mass spectrometry (MS) to identify proteins or peptides separated via liquid phase separation techniques and/or gel-based separation techniques have led to significant technological advance in protein and peptide expression analysis. There are two main methods for the mass spectrometric characterization of proteins and peptides: matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI). Using various approaches, MALDI and ESI ion sources can be combined with time-of-flight (TOF) or other types of mass spectrometric analyzers to determine the masses or the sequences of peptides.

In MALDI, peptides are co-crystallized with the matrix, and pulsed with lasers. This treatment vaporizes and ionizes the peptides. The molecular weights (masses) of the charged peptides are then determined in a TOF analyzer. In this device, an electric field accelerates the charged molecules toward a detector, and the differences in the length of time it takes ionized peptides to reach the detector (their time-of-flight) reveal the molecular weights of the peptides; smaller peptides reach the detector more quickly. This method generates mass profiles of the peptide mixtures - that is, profiles of the molecular

weights and amounts of peptides in the mixture. These profiles can then be used to identify known proteins from protein sequence databases.

In ESI and a technique called liquid chromatography (LC)/MS/MS, a voltage is applied to a very fine needle that contains a peptide mixture. The needle then sprays droplets into a mass spectrometric analyzer where the droplets evaporate and peptide ions are released. In LC/MS/MS, researchers use microcapilliary LC devices to initially separate peptides.

Mass spectrometry (MS) is a valuable analytical technique because it measures an intrinsic property of a bio-molecule, its mass, with very high sensitivity. MS can therefore be used to measure a wide range of molecule types (proteins, peptide, or any other bio-molecules) and a wide range of sample types/biological materials. Correct sample preparation is known to be crucial for the MS signal generation and spectra resolution and sensitivity. Sample preparation is therefore a crucial area for overall feasibility and sensitivity of analysis.

Proteins occur naturally within cells, as components of cellular structures and as components of natural biological fluids such as blood, urine, saliva, tears, lymph and sweat. Proteomics is an essential tool for studying biological systems and processes because proteins provide a rich source of valuable information. For example, this information allows the comparison of biomarkers that may differ qualitatively and/or quantitatively between healthy and diseased population groups. Proteomics technologies allow the identification of individual protein species in complex mixtures of proteins.

Proteomics are being used in drug discovery and development, for example to detect proteins significantly altered in patients with particular conditions or diseases. Some of these disease-associated proteins may be identified as novel drug targets and some may be useful as biomarkers of disease progression. Such biomarkers may be used to improve clinical development of a new drug or to develop new diagnostics for the particular disease.

Detection of disease-associated proteins may be achieved by the following method. Protein samples are taken from both diseased individuals and healthy individuals. These samples may be cells, tissues, or biological fluids that are processed to extract and enrich protein constituents either in solution or as solid material. After high-throughput analysis (proteomics), protein expression fingerprints are produced for diseased and for healthy individuals. These two types of fingerprint are combined in a computer database and analysed using bioinformatic tools to identify and select the proteins whose expression is affected by the disease. These proteins are then further characterised and detailed profiles are produced.

Various disease-associated proteins are known, and some of these are enzymes whose activity increases or decreases at some stage in the development of a particular condition or disease. Such enzymes may be suitable drug targets, leading to a search for pharmaceutically-active compounds (drugs) that could be used to inhibit or stimulate the enzyme and thus prevent or treat the condition or disease. Other disease-associated proteins may be degradation products of particular enzymes, or proteins that are made more abundantly in the presence of the disease.

Examples of disease-associated proteins include those enzymes that have been implicated in the onset and/or progression of Chronic Obstructive Pulmonary Disease (COPD), as discussed below.

COPD, which is mainly caused by cigarette smoking, is expected to be the third leading cause of death worldwide by the year 2020. COPD is characterised by reduced maximum expiratory flow and slow forced emptying of the lungs. These airflow limitations are mainly due to chronic bronchitis, involving hypertrophy of mucous glands, and emphysema produced by destruction of alveolar walls. The latter leads to enlargement of the air spaces distal to the terminal bronchiole, with consequent collapse of small airways, limitations of the airflow, destruction of parts of the capillary bed, and loss of the elastic

recoil of the lung. This loss of elastic recoil and the enlargement of the air spaces in the lungs of COPD patients lead to reduced values of forced expiratory volume (FEV), and increased values of forced vital capacity (FVC). Disease severity is determined as the degree of lung function impairment, which is measured with a spirometer. The presence of a postbronchodilator $FEV_1 < 80\%$ of the predicted value in combination with an $FEV_1/FVC < 70\%$ confirms the presence of airflow limitation that is not fully reversible. The chronic exposure to cigarette smoke causes an inflammatory response in the lung, leading to changes in the airway epithelial surface and to activation and an increased number of several inflammatory cells.

Inflammation and a protease-antiprotease imbalance have long been proposed to act as downstream effectors of the lung destruction following chronic cigarette smoking. Histological studies have demonstrated increased numbers of macrophages and T-lymphocytes in the airways of smokers, and also an increase of neutrophils in the airways of smokers and COPD patients, which related to the severity of the airway obstruction. Alveolar macrophages are long-lived phagocytes, and are the most abundant defence cells in the lung both under normal conditions and during chronic inflammation. By sending out chemotactic factors they then recruit neutrophils and lymphocytes by activating adhesion molecule expression on pulmonary microvascular endothelial cells at the site of infection. The inflammatory cells invading the smoker's lung produce mediators locally, such as cytokines, neutrophil elastase, serine- and metalloproteases, and oxidants. These mediators, which likely play an important role in the development of COPD, can act to further activate the immunoresponse, and also to degrade the components of the extracellular matrix.

Normally plasma proteinase inhibitors, especially α_1-antitrypsin (α_1-AT), prevent proteolytic enzymes like neutrophil elastase (NE) from digesting structural proteins of the lung. According to the proteinase-antiproteinase hypothesis, emphysema result from an increase of proteinase release in the lungs, a reduction in the antiproteinase defence, or a combination. NE cleaves elastin, which is the principal component of elastic fibres

constituting a main part of the lung's extracellular matrix. Studies show that individuals who are homozygous for α_1-AT deficiency have an increased susceptibility for developing pulmonary emphysema, especially if they also smoke. Apart from cleaving elastin NE also has a role in the recruitment of neutrophils. It stimulates macrophages to release leukotrine B_4 (LTB_4), a chemoattractant of neutrophils in the peripheral airways, which leads to an amplification of neutrophil recruitment believed to cause the development of early emphysema. NE also stimulates the secretion of mucous and the release of another neutrophil chemoattractant, IL-8, from epithelial cells, which perpetuates the disease process. Further more NE causes loss of ciliated epithelium, reduces ciliary beat frequency, and impairs other host defences, which may in part be responsible for bacterial colonisation.

One example of a disease-associated enzyme linked to COPD is the matrix metalloproteinase MMP12, also known as macrophage elastase or metalloelastase. MMP12's substrate is elastin, the insoluble, elastic protein of high tensile strength found in intercellular spaces of the connective tissues of large arteries, trachea, bronchi and ligaments. MMP12 was initially cloned in the mouse by Shapiro *et al* [1992, Journal of Biological Chemistry <u>267</u>: 4664] and in man by the same group in 1995. MMP-12 is preferentially expressed in activated macrophages, and has been shown to be secreted from alveolar macrophages from smokers [Shapiro *et al*, 1993, Journal of Biological Chemistry, <u>268</u>: 23824] as well as in foam cells in atherosclerotic lesions [Matsumoto *et al*, 1998, Am J Pathol <u>153</u>: 109]. A mouse model of chronic obstructive lung disease (COPD) is based on challenge of mice with cigarette smoke for six months, two cigarettes a day six days a week. Wildtype mice developed pulmonary emphysema after this treatment. When MMP12 knock-out mice were tested in this model they developed no significant emphysema, strongly indicating that MMP-12 is a key enzyme in the COPD pathogenesis. MMP12 is believed to degrade lung tissue by degrading the elastin within the tissue. The role of MMPs such as MMP12 in COPD (emphysema and bronchitis) is discussed in Anderson and Shinagawa, 1999, Current Opinion in Anti-inflammatory and Immunomodulatory Investigational Drugs <u>1(1)</u>: 29-38. It was recently discovered that

smoking increases macrophage infiltration and macrophage-derived MMP-12 expression in human carotid artery plaques Kangavari [Matetzky S, Fishbein MC *et al.*, Circulation 102:(18), 36-39 Suppl. S, Oct 31, 2000].

The current use of proteomics or peptidomics in drug discovery and development (particularly for the disease COPD) is limited by various factors, including for example:

a) the lack of profiles of disease-associated peptides that can be linked to *specific* drug targets (because current fingerprinting methods analyse total protein differences, and do not focus on a particular protein/drug target);
b) the lack of biomarkers to identify COPD sufferers at an early stage of the disease;
c) the lack of biomarkers to evaluate potential drugs that are MMP12 inhibitor compounds, particularly in clinical studies (ie for validation that the MMP12 target is hit by the inhibitor).

We have now developed a new methodology for producing and using protein/peptide fingerprints, allowing us to identify and investigate disease-associated proteins/peptides that can be linked to specific drug targets (such as MMP12), or to specific drug target combinations.

DESCRIPTION OF THE INVENTION

We now provide a new method, composed of multiple linked steps, for detecting the naturally occurring products of protein expression and protein catabolism. This method may be used for biomedical evaluation and biomedical characterization.

The method may be applied to naturally occurring substances or mixtures of naturally occurring substances and synthetic substances. The method results in the identification of specific protein substituents, also known as peptides, present as separate entities or present within complex mixtures of proteins and peptides. Each peptide may be defined by a specific sequence of amino acids in alignment.

Ink. t. Patent- och reg.verket

2003 -09- 2 5

The method may measure either the presence or absence of the peptides within complex mixtures of proteins and peptides. The method may be applied for the biomedical study of the relationships between the expression and function of proteolytic enzymes and the status of protein degradation products, hereafter referred to as unique peptides.

The method allows the identification of some or all peptides which are proteolytic breakdown products of a given enzyme with a given substrate and which are measurable (or example using MS identification).

The method combines several key steps together which results in the specific separation, isolation, and identification of unique peptides present in biological material. The unique peptides are the constituent units of protein molecules identifiable for example by MS or other methodologies.

The method may be applied to human clinical samples.

We provide a multi-step method for identifying:
1) the unique peptide identity presented for example as atomic mass units of entities resulting from the proteolytic interaction of a given proteolytic-enzyme with its given substrate;
2) the preparation, separation, and identification of peptides derived from given proteolytic enzymes with given subtrates either occurring naturally or produced in a laboratory setting;
3) the presence and/or absence of these same exact peptides in biological samples, for example within a sample of human blood or urine;
4) the statistical method for determining the identification of either naturally occurring peptides partitioned into fractions by the separation methodology referred to in 1) to 3) above;
5) the statistical method for defining the presence and/or absence of these same peptides in multiple human subjects, collected and grouped by, for example clinical disease status.

The method may be used to determine whether certain proteolytic enzyme processes are occurring or have occurred in human subjects. This may allow association of the presence

and/or absence of certain products of proteolytic digestion, for example peptide fingerprints, in certain persons, or persons with known diseases, or persons with known stages or phases of disease. The method may allow measurement of the presence or absence or quantity of specific peptide fingerprints within human clinical samples such as for example urine or blood.

The method may allow monitoring of the effect of certain and or all medicines or substances which effect the expression or function of proteolytic enzymes. The method may allow us to measure the presence or absence or quantuity of specific peptide fingerprints within human clinical samples such as for example urine or blood as a result of medical and or pharmacological intervention.

The method first identifies all or some of the peptides produced by a given enzyme with a given substrate in a controlled laboratory setting. This step optimizes the chances for producing all likely candidate peptide fragmentts from a given enzyme-substrate reaction. This includes attention to Michalis Minton kinetics for maximizing the ratio of reactants, the pH level and salt concentrations used in the reactant solutions, the temperature of the reaction, the time of the reaction, for example. This may result in the production of stable end form unit length entities of unique peptides which are present in the reactant solution [Reaction product 1]. The net result of the optimised laboratory controlled reaction of the given enzyme with the given substrate is a signature peptide profile for that reaction. This collection of unit length peptides resulting from proteolytic digestion is also referred to as a peptide fingerprint. Scheme A (below) illustrates this step (generation and identification of peptide entities, peptide fingerprint, and unique peptide atomic mass identification as products of laboratory controlled digestion of human protein substrates with given protein enzyme):

DEFINED REACTION				REACTION PRODUCT 1		REACTION PRODUCT 2
controlled reaction						
protein enzyme	**+**	**human protein**	**=**	**unique peptide fingerprint**	**→**	**isolation and identification of unique peptide**
optimised conditions						

Ink. t. Patent- och reg.verket

2003 -09- 25

The peptide fingerprint is then subjected to a series of biochemical separation steps (described below) to fractionate the individual unique peptides by their intrinsic bio-physical properties, for example charge, size, and hydrophobicity properties. The individual fractions of the unique peptide fingerprint are then identified using MALDI MS to determine precise atomic mass measurements for each unique peptide entity. The net result of this fractionation and identification process is a quantitative and qualitative list of all peptide fragments produced and comprising the peptide fingerprint [Reaction product 2]. This list of atomic mass identities is then used in further steps of the method according to the invention.

In a first aspect of the invention, we provide a method to generate a peptide fingerprint of the degradation products of a disease-associated enzyme X, wherein enzyme X is associated with disease Y, which comprises:

(a) mixing a disease-associated enzyme X with its natural substrate *in vitro* in conditions that allow interaction between enzyme X and its substrate;
(b) allowing the substrate to be degraded by enzyme X;
(c) analysing the mixture to produce a peptide fingerprint of the degradation products.

The peptide fingerprint produced by the method of the invention may be used in the diagnosis or study of disease Y, for example to aid the discovery and development and administration of drugs to treat disease Y, particularly drugs wherein enzyme X is the drug target.

Disease Y is any condition or disease affecting humans or non-human animals. In particular, disease Y is any condition or disease affecting humans. For example, disease Y may be a condition or disease affecting the respiratory tract (such as COPD), the cardiovascular system, the gastrointestinal tract, the neurological system, the endocrinological system, the immunological system. In addition, disease Y may be an

allergic condition or disease, an infectious condition or disease, or an oncological condition or disease.

The disease-associated enzyme X is any enzyme that shows increased activity during the onset or progression of any condition or disease affecting humans and non-human animals (particularly humans). This increased activity causes or contributes to disease onset or progression. The disease-associated enzyme X may be a drug target.

The degradation products in the mixture will be peptides generated by breakdown of the substrate by the enzyme. The mixture is analysed by peptidomics technologies to produce a peptide fingerprint (a constant, reproducible set of the degradation products). The peptide fingerprint and/or selected peptides or groups of peptides may be useful as biomarkers relating to disease Y (including its presence, its development and/or its treatment).

To determine the entities within the peptide fingerprint and produce Reaction Product 2 (see scheme A above), the peptides produced as Reaction Product 1 are first separated (for example, by chromatographic separation, liquid phase separations) utilising mechanisms such as:

A) size exclusion (in samples where fractionation is required based upon size);

B) hydrophobic interactions (utilisation of reversed phase separation mechanisms whereby peptides will be separated by hydrophobicity);

C) polar interactions (silanol and other types of polar functionalities readily interact with polar peptides and can be separated due to polar chromatographic interactions);

D) chiral affinity (chiral small molecules may be used as selective ligands for peptide binding and thus separation);

E) metal affinity (chelation by metal ion interaction of amine, and or carboxy-hydroxy functional groups, as well as Nickel ion-Histidine peptide residues, iron-, Gallium-ions and phosphate functionalities on peptides);

F) antibody binding (traditional antibody-antigen immunoaffinity bindings with both weak-medium-strong affinities, with binding constants ranging from 10^5 to 10^{10}).

After separation, the peptides are profiled by ascertaining their physiochemical properties plus accurate masses (the peptide index, comprising the size, polarity/charge and hydrophobicity of the peptides). This is optionally followed by sequencing of the peptides.

Scheme B below illustrates the technology platform for analysis of unique peptide fingerprints resulting in individual mass identities of peptide entities:

REACTION PRODUCT 1		REACTION PRODUCT 2
unique peptide fingerprint	→ *multi-step process* → *isolation & identification of unique peptide entities*	**unique peptide entity identities**

The method according to the first aspect of the invention may be used to identify biomarkers for a particular disease Y that is known to be associated with a particular drug target (enzyme X). The peptide fingerprint of the degradation products is used as a biomarker for disease Y.

In a preferred method according to the first aspect of the invention, enzyme X is MMP12 and disease Y is COPD. As an example of a method according to the first aspect of the invention, monocyte elastase (comprised partly or wholly of MMP12) is mixed with human elastin. In this method, the disease-associated enzyme X is MMP12 which has the natural substrate elastin and is associated with COPD. Conditions are optimised to ensure high MMP12 activity and good degradation of elastin. Michaelis-Menten kinetics are used to determine the preferred stoichiometry of reactants, and the substrate type and amount are chosen to give a favourable equilibrium constant for the progress of the reaction.

In a second aspect of the invention, we provide a method to determine if or confirm that an enzyme X is associated with disease Y which comprises:

Ink. t. Patent- och reg.verket

2003 -09- 25

(a) obtaining a healthy biofluid sample or a healthy tissue sample;
(b) analysing the healthy sample to produce a healthy peptide fingerprint;
(c) obtaining a diseased biofluid sample or a diseased tissue sample, wherein the diseased sample shows signs of the onset or progression of disease Y;
(d) analysing the diseased sample to produce a diseased peptide fingerprint;
(e) comparing the healthy peptide fingerprint to the diseased peptide fingerprint and identifying the set of peptides found only in the diseased peptide fingerprint;
(f) mixing enzyme X with its natural substrate *in vitro* in conditions that allow interaction between enzyme X and its substrate;
(g) allowing the substrate to be degraded by enzyme X;
(h) analysing the mixture to produce a peptide fingerprint of the degradation products;
(i) comparing the diseased set of peptides identified in step (e) with the peptide fingerprint of the degradation products produced in step (h), and determining if there are statistically significant similarities between them;
(j) if there are statistically significant similarities between the diseased set of peptides identified in step (e) and the peptide fingerprint of the degradation products produced in step (h), concluding that enzyme X is associated with disease Y;
(k) if there are no statistically significant similarities between the diseased set of peptides identified in step (e) and the peptide fingerprint of the degradation products produced in step (h), concluding that enzyme X is not associated with disease Y.

In a method according to the second aspect of the invention, human clinical material is analysed using the methodology described below. The quality of the human clinical material is an important factor in obtaining accurate measurements. The methods for accurately determining, identifying, or measuring unique peptide entities in human clinical material are direcly dependent upon certain criteria. Quality is defined in relation to human clinical material as follows. Clinical samples should be obtained using methods which preserve the integrity of proteins in a natural state, and minimize the effects of denaturation, and destruction. This includes careful sample preparation, and storage under

conditions which preserve protein structure and function. Human clinical material should be well documented in the features of clinical presentation which these samples represent. Information which relates the sample to specific aspects of the disese such as the clinical presentation of disease, for example stages or phases of disease, or noted impairments of structure and function characteristic or not of these diseases. Samples from subjects should be identifiable for example as being free or not free from obvious diseases. When possible the best practice should be the linkage of disease with the individual samples, and with other subject samples with similar linkages to disease. When possible the best practice should be the linkage of a particular site or location of disease with the individual samples

When possible the best practice is to obtain as much information regarding the sample, the history of the sample, and the medical classification of the sample as possible. It is also important to obtain as much information as possible regarding the phases of disease reflected in indivual samples.

Scheme C below illustrates the preferred description of quality human clinical samples:

QUALITY CLINICAL SAMPLES	→	*Critical sampling* *Processing* *Storage analysis* -------------------- *Subject grouping* *Clinical classification* *Clinical information* *Relate sample to disease site*	→	*Statistical grouping* *Relating sample result to clinical presentation*

The biofluid or tissue sample may be derived from any part of the human or non-animal body (including cells grown *in vitro*), preferably from any part of the human body. For example, the sample may be derived from urine, blood, sputum, saliva, nasal secretions, exhaled breath condensate, bronchoalveolar fluid, bronchial fluid or any other biological fluid or tissue. A tissue sample is defined as a sample comprising one or more cells. A biofluid is defined as any sample of clinical material in solution form (preferably human clinical material). This may include blood, serum, plasma, saliva, lavages, tears, urine,

Ink. t. Patent- och reg.verket

2003 -09- 2 5

seminal fluid, joint fluid, aqueous humor, washings of cavities or sinuses, the soluble form of tissue preparations, the soluble form of organ preparations, or sweat. The samples may be derived from singular subjects or pools of singular samples from multiple subjects.

As defined herein, healthy biofluid or tissue samples are samples from individuals without recognised clinical disease or symptoms of disease. Healthy biofluid or tissue samples may represent the average or normal variation of expression of gene products in the human population that do not show any signs of disease onset or progression. As defined herein, diseased biofluid or tissue samples are samples from specific identified individuals that have been clinically evaluated and diagnosed for specific disease processes, or who show symptoms of clinical disease which are not yet categorised clinically as a specific disease. Diseased biofluid or tissue samples may express qualitatively and/or quantitatively different sets of peptides/proteins/endogenous products from healthy biofluid or tissue samples. Such differences include changes in the steady state, changes in destructive processes present in resident and non-resident cells, changes in differentiation states and changes in repair processes. Differentiation states are defined as stages of maturity in cell function and/or phenotype.

The biofluid or tissue samples are obtained by acquiring, processing and preparing the biological material. The methods according to the invention may be used for both small scale and large scale clinical investigation, with typical patient groups of 10-20 patients in each study group. The clinical study material needs to be of high quality and this is ensured by optimised sampling, sample handling, and sample storage protocols. These sample protocols ensure the minimum degradation of the naturally occurring proteins and peptides present within these samples.

The biofluid or tissue samples and the enzyme/substrate mixture are analysed by peptidomics technologies to produce peptide fingerprints. These peptide fingerprints are obtained as explained above for the method according to the first aspect of the invention

(peptide separation by various mechanisms followed by determination of physiochemical properties plus accurate masses, optionally followed by sequencing of the peptides).

When comparing the diseased set of peptides identified in step (e) with the peptide fingerprint of the degradation products produced in step (h), it is necessary to determine if there are statistically significant similarities between them.

The diseased peptide profile may differ depending on the disease stage of the diseased biofluid or tissue sample. It is possible to generate peptide profiles from samples showing early stage disease, mild disease, moderate disease or severe disease. Thus pathology of a disease may be linked to peptide profiling.

The diseased peptide profile may also differ depending on the source of the diseased biofluid or tissue sample. It is possible to generate peptide profiles from samples taken from different compartments of the human or non-animal body.

The diseased peptide profile may also differ depending on the individual human or non-human animal from which the sample was derived, or on the particular group to which the human or non-human animal belongs.

In a variation of the method according to the second aspect of the invention, multiple biofluid or tissue samples are used wherein each diseased sample has the same disease.

In another variation of the method according to the second aspect of the invention, multiple disease sets are used (by using more than one sample each having a different disease, or by using one sample having more than one disease).

Combined analysis of diseased peptide profiles may be used to reduce a multifactorial disease processes to its component parts. Each part and its relation to other parts may be analysed.

In a preferred method according to the second aspect of the invention, enzyme X is MMP12 and disease Y is COPD.

In a third aspect of the invention, we provide a method to determine the presence of a disease Y which comprises:

(a) obtaining a biofluid sample or a tissue sample;

(b) analysing the sample to obtain its peptide fingerprint;

(c) mixing enzyme X with its natural substrate *in vitro* in conditions that allow interaction between enzyme X and its substrate, wherein enzyme X is associated with disease Y;

(d) allowing the substrate to be degraded by enzyme X;

(e) analysing the mixture to produce a peptide fingerprint of the degradation products;

(f) comparing the peptide fingerprint of the sample identified in step (b) with the peptide fingerprint of the degradation products produced in step (e), and determining if there are statistically significant similarities between them;

(g) if there are statistically significant similarities between the peptide fingerprint of the sample identified in step (b) and the peptide fingerprint of the degradation products produced in step (e), concluding that disease Y is present;

(h) if there are no statistically significant similarities between the peptide fingerprint of the sample identified in step (b) and the peptide fingerprint of the degradation products produced in step (e), concluding that disease Y is absent or is being successfully treated.

Statistically significant similarities may be single peptide or multi-peptide identity. Determining statistically significant similarities involves using a degradative product peptide fingerprint (for example, an MMP12/elastin degradative product peptide fingerprint) to search for comparative fingerprints within biofluid or tissue samples.

In a preferred method according to the third aspect of the invention, enzyme X is MMP12 and disease Y is COPD.

The method according to the third aspect of the invention may be used to determine the presence of a disease Y in humans or in non-human animals. For example, the method may be used during clinical trials involving individual humans or in pre-clinical trials involving non-human animal models. The humans or non-human animals may appear to be healthy or may appear to be diseased. Those that appear to be healthy may be healthy or may be clinically asymptomatic subjects.

In a fourth aspect of the invention we provide a diagnostic test kit for determining the presence of a disease Y which comprises means to compare the peptide fingerprint of a biofluid sample or the peptide fingerprint of a tissue sample with the peptide fingerprint of the degradation products in a mixture of enzyme X with its natural substrate, wherein enzyme X is associated with disease Y. Preferably enzyme X is MMP12 (natural substrate is elastin) and disease Y is COPD.

In a fifth aspect of the invention we provide a method to analyse the effect of a drug Z on enzyme X, wherein enzyme X is associated with disease Y, which comprises:

(a) treating a human or non-human animal with the drug Z, wherein the human or non-human animal is suffering from disease Y;
(b) obtaining a biofluid sample or a tissue sample from the human or non-human animal;
(c) analysing the sample to obtain its peptide fingerprint;
(d) mixing enzyme X with its natural substrate *in vitro* in conditions that allow interaction between enzyme X and its substrate, allowing the substrate to be degraded by enzyme X;
(e) analysing the mixture to produce a peptide fingerprint of the degradation products;

(f) comparing the peptide fingerprint of the sample identified in step (c) with the peptide fingerprint of the degradation products produced in step (e), and determining if there are statistically significant similarities between them;

(g) if there are statistically significant similarities between the peptide fingerprint of the sample identified in step (c) and the peptide fingerprint of the degradation products produced in step (e), concluding that drug Z is not inhibiting enzyme X;

(h) if there are no statistically significant similarities between the peptide fingerprint of the sample identified in step (c) and the peptide fingerprint of the degradation products produced in step (e), concluding that drug Z is inhibiting enzyme X.

The method according to the fifth aspect of the invention may be used during drug discovery and development to ascertain whether the correct drug target is being affected when treating with a particular drug Z. Drug Z may be a drug or a candidate drug compound. The method allows direct study of the effect of drug Z on enzyme X, including the effect of different levels of drug Z. The peptide fingerprint of the degradation products in a mixture of enzyme X with its natural substrate is a biomarker. Use of this biomarker improves the efficiency of clinical trials of drug Z, by allowing use of smaller patient groups giving cost and time savings.

In a preferred method according to the fifth aspect of the invention, enzyme X is MMP12 and disease Y is COPD.

In a sixth aspect of the invention we provide a diagnostic test kit for analysing the effect of a drug Z on enzyme X which comprises means to compare the peptide fingerprint of a biofluid sample or the peptide fingerprint of a tissue sample with the peptide fingerprint of the degradation products in a mixture of enzyme X with its natural substrate, wherein the sample has been obtained from a human or non-human animal that has been or is being treated with the drug Z. Preferably enzyme X is MMP12.

From the methods according to the invention, it is possible to generate a disease model (a predictive indicator of disease development), by analysing biofluid or tissue samples over time and looking for changes in the peptide fingerprint.

In all the methods according to the invention, it is preferable to use the following methodology to generate the protein/peptide fingerprints. This methodology gives the optimal resolution and sensitivity.

The preferred methodology is an automated multidimensional liquid phase separation platform technology. The entire platform is operated automatically in a closed operation system, where the multidimensional separating mechanisms are performed in liquid separation phases on chromatographic columns. The interconnections of these separation steps are performed on-line with transfer steps in-between the columns within the workstation. The interfacing in-between the separation mechanisms is provided by chromatographic conditions that allow the analytes to be transferred from one dimension to the next without losses. This is accomplished by the liquid-liquid transfer in-between the dimensions. An operational description of the methodology is given below.

Biofluid samples are introduced into the liquid phase peptide profiling platform and kept at 4°C thermo stated to ensure stability of the samples over time. The sample is then injected into the first dimensional separation from the autoinjector (column 1). The mechanism in this step is based upon size separation (the separation packing material, of polymer or silica origin, has highly defined pores). In the first dimension, larger sized proteins and biopolymers will be excluded from entering the pores of the beads of the separation material. The analytes of interest, such as peptide analytes, diffuse into the pores and bind to the functionality within the pores. This functionality can be electrostatic charged surfaces or hydrophobic surfaces onto which the peptides are bound. In this way selective enrichment of the peptides occurs as simultaneously the larger sized proteins and biopolymers are excluded and eluted to waste. The column material is then washed a few times with varying eluents in order to exclude interfering components from the sample that

has bound to the outer surface as well as to filters and exposed surfaces of the chromatographic system. In this way, the enriched peptide fraction in the pores of the column material is isolated with a high purity.

After the washing steps, a strong eluent is introduced into the next dimension, column 2. In the second dimension, the elution from column 1 is transferred into column 2 on-line and adsorbed on top of the support of column 2. Column 2 is a bead with charged functionality where the separation is performed by electrostatic mechanisms. A gradient elution is used, and the corresponding peptides are separated in eluted fractions. These fractions are next separated in the third dimension by hydrophobicity, whereby the salt is eliminated from the peptide fractions and concentrated using a washing step of aqueous media, followed by elution onto a target plate surface from where the peptide mass sequences is determined.

The fourth dimension of the system utilizes mass spectrometry where the mass, intensity (quantity) of each and every peptide component of all the fractions of the sample is analyzed.

The data generated from the Mass Spectrometer comes as files with three headings: Fraction, Mass/z and Intensity. Since masses have a large number of decimal values, these are rounded off to the nearest digital value. A data matrix is contructed from the Fraction, mass/z and intensity columns, so that the row names of the new matrix are given by the combinations of fractions and digitized masses present in the data, and where the intensities are summed for each such mass and fraction combination. The data matrix then becomes

	Subject1	Subject2 ....
Fraction1 x mass1	$I_{1,1}$	$I_{1,2}$
Fraction2 x mass2	....	....

where $I_{1,1}$ etc are the summed intenisties. Subjects are normalised by equating the total sum of intensities per subject.

A Java code calculates a regularized *t*-statistic that minimizes the false positive and false negative rates, following the theory in Broberg (2003, *Genome Biology*, 4(6):R41). In practice one starts out with a top list size or a number of practical top list sizes, and the task is to find an optimal size in the range given and to populate that list with as many true positives as possible. The test statistic used has the form pioneered by Tusher *et al* (2001, *Proc. Natl. Acad. Sci.USA*, **98:** 5116-5121)

$$d = \frac{diff}{S_0 + S}$$

where *diff* is an effect estimate, e.g. a group mean difference, S is a standard error, and S_0 is a regularizing constant. In the two sample case putting $S_0=0$ will yield the equal variance *t*-test. Using estimates of the false positive and false negative rates an optimisation procedure minimises the criterion $C = \sqrt{(FP^2 + FN^2)}$ over a lattice of possible values of S_0 (given by percentiles in the distribution of S) and the length of the top list.

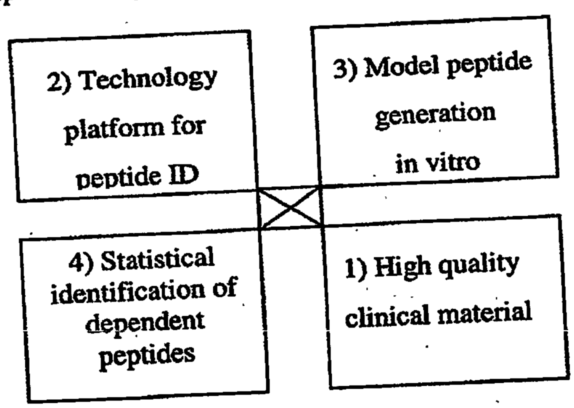
The output include group means, p-values for the comparison, the false positive rate and false negative rate that would arise from including the current Fraction x Mass and all with smaller p-values. The cut-off is chosen as to minimise the false positive and false negative rates.

The resulting mass spectra data, peptide mass, peptide fraction and peptide identity is generated by statistical comparisons between COPD and healthy subjects. The digital part of the mass is grouped into bins with +/- 0.5 mass units on either side of the detected mass, and combined with a given peptide fracion. Next the bin intensities are summed to produce an extrapolated identity for each and every fragment. The total bin numbers used in the statistical analysis were typically between 10.000-20.000. The mass fragments are then compared by subject groupings such as COPD or healthy. The statistical analysis is based on 40-50 fractions collected from each subject. The cycle time generating the 40-50 peptide fractions is less than 5 hours.

The mass fragments of the peptides described in Example 1 and Example 2 below could not be identified in any of the 20 healthy subjects tested using this methodology.

Integrated process steps for biomarker identification is essential, containing the following four process defining corner stones; 1/ high quality biomedical clinincal material, 2/ Technology platform for qualitative and quantitative determination of peptides, 3/ In Vitro assay where qualitatigve and quantitative analysis of resulting MMP-12 enzyme products determined using human Elastin as the substrate, 4/ A statistical method for relating of multiple sets of peak identities to presents or differential expression within designated subbject groups. This is preferred way of analysing the data that allows to analyse the biomarker peptides that are likely associated with COPD patient urine (Scheme D). Each of the four process corner stones are interdependent and required in order to determine biomarker peptides by differential quantitation between healthy and COPD subjects, relating it to the MMP-12 enzyme function /activity derived peptides from Elastin.

Scheme D: interdependence of process cornerstones



The invention is illustrated by the following non-limiting examples.

EXAMPLE 1

The peptide fragments of elastin, which are the result from the enzymatic digestion of elastin by MMP 12, are first identified under laboratory conditions using a model in vitro system for generating unit length peptides and peptide fingerprints of the proteolysis of elastin by this MMP12 enzyme. These peptides are then collected and fractionated on a series of chromatography columns in order to isolate the peptides for identification. The peptides are then identified as unit atomic mass entities using MALDI-TOF mass spectrometry spectra. For each fraction of column separated peptides, the mass of each entity, and the mass intensity of each entity are recorded and stored into databases. A statistical software program is used to analyze the profile of atomic mass unit identities from all fractions analysed by MS. A concensus atomic mass identity is assigned for a set of landmark elastin peptides associated with MMP12 enzymatic degradation. This set of elastin derived peptides also known as a peptide fingerprint is shown in Table 1. This part of the invention provides the identities of landmark peptides which can be used as reference points for discovering and or identifying peptides with similar or nearly similar physical-chemical properties in other complex mixtures of proteins. This part of the invention provides the identities of landmark peptides which can be used as reference points for discovering and or identifying peptides of similar characteristic that are present in human clinical samples. This part of the invention provides the identities of landmark peptides which can be used as reference points for discovering and or identifying the exact amino acid sequence identity of these same peptide entities.

Experimental procedure

The following method was used. An MMP-12 *in vitro* assay has been developed in order to make peptide annotations that are directly assigned to the MMP-12 activity. These peptide masses are unique to MMP-12 and is used to match the presence of these peptides within biofluid samples sampled from both healthy and diseased patients. The assay is run as follows; human lung Elastin is used as the substrate in the in vitro assay with human MMP-12. The insoluble human Elastin is washed using a 100 mM TRIS-HCL buffer pH 7.5 containing 0. 1 M NaCl and 10 mM CaCl2, and then centrifugated in-between

repeated washing steps. Next, 1.2 mg Elastin is re-suspended in the assay buffer 100 mM TRIS-HCL buffer pH 7.5 containing 0. 1 M NaCl and 10 mM CaCl2 and 60 µg human MMP-12. Incubation was made at 37°C for 7 hours where the elastin was degraded by the enzyme MMP12. The proteolysis process was stopped by the addition of iodoacetamide. After digestion, the samples were analysed directly or kept at -80°C in the freezer. The samples were analysed by thawing of samples in room temperature, and a sample preparation step was performed using a reversed phase preparation step. The sample was then eluted from the preparation by an acetonitrile elution step onto the MALDI-TOF target plate. Cyano-4-hydroxycinnamic acid (ACHA) was added as the matrix for crystal formation and run on the MALDI-TOF mass spectrometer where a peptide fingerprint of the MMP12/elastin degradation products;were identified and annotated.

The specific experimental conditions used in order to generate differentially displayed peptides in human urine samples from healthy and COPD patients was made as follows; The biofluid was sampled from patients and thereafter aliquoted and frozen at -80°C. The frozen urine was thawed in room temperature, pH adjusted to 2.5 with ortophosphoric acid and processed for HPLC separations.The urine samples were introduced into a two-dimensional chromatography system where the first separation mechnism utilised was size exclusion chromatography. The cut-off of the column material was around 15 kDa. The fractionations resulting from the size exclusion separation step is next transferred on-line to a cation –exchange step where the peptides/proteins will be separated based upon charge. These fractions are then transferred to a reversed phase separation step where all interfering matrix components present in the sample is eliminated, this is the third dimensional separation. The third dimension fractions are the spotted down onto a MALDI target plate by a robotic feeder that adds the MALDI matrx to the peptide/årootein sample sptos. Next the MALDI sample paltae is inserted in the MALDI-TOF mass spectrometer instrument and analysed according the exact mass, quantity and isotope resolution. Figure 5 and Figure 6 illustrates two examples where we have qualitative and quantitative differences in-between helathy and COPD patent data from resulting Elsatin peptides in the urine samples, respectively. In Figs 5A and 6A, the biomarker peptides are found present

Ink. t. Patent- och reg.verket

2003 -09- 2 5

in the biofluid from COPD patients, due to active MMP-12 degradation, while Figs 5B and 6B does not hold these peptides. The absense of these peptides in the healthy subjects were not possible to quantitate (found absent) at the lowest level of quantification that MALDI instrumentation offers.

The elastin peptide fingerprint generated by MMP12 digestion under laboratory conditions, and described above, is used as a reference landmark for finding identical or nearly identical homologous peptides within clinical biofluids.

Elastin peptide fragments or the elastin peptide fingerprint are identified in the urine of a patient with Chronic obstructive Pulmonary Disease (COPD). This patient has been previously identified within a clinical setting as having COPD. In this example the patient showed abnormal airway function tests, as revealed by a low FEV1 score. This patient further showed evidence of pulmonary alveolar hyper-inflation and emphysema using CT imaging. This patient further showed evidence of elastin protein destruction within the parenchyma of the lung by histology examination of the lung, and by using a method for identifying elastin protein within lung tissue by immunohistochemistry with an antibody specific for human elastin, and specific for a hexamer epitope of human elastin and tropoelastin. This patient further showed histological evidence of alveolitis and alveolar macrophage accumulation in areas of lung, located adjacent to elastin expression and elastin degradation. This patient further showed histological evidence that the same alveolar located macrophages were activated to express a marker of activation, CD-68, within tissue by immunohistochemistry of sections of lung tissue with an antibody specific for CD-68. This patient further showed histological evidence that the same alveolar located macrophages were activated to express human MMP-12 within tissue by immunohistochemistry of sections of lung tissue with an antibody specific for human MMP-12. This patient was a member of a group of 20 patients under study. The findings described above were not unique to this single COPD patient. The findings described above were a common finding among the 20 study COPD patients.

The following method was used:

1) a healthy human bio fluid sample was obtained; We define healthy in this example as a living adult person aged 20-80, without symptoms of disease, without current clinical disease, not being treated for a clinical disease with medication or prescribed drugs, without at risk behaviour for developing disease such as smoking, drug abuse, alcoholism, overweight, or without a diagnosed genetic disposition for disease in later onset of life for example. We define biofluid here as any sample of human clinical material in solution form. This may include blood, serum, plasma, saliva, lavages, tears, urine, seminal fluid, joint fluid, aqueous humor, washings of cavities or sinuses, the soluble form of tissue preparations, the soluble form of organ preparations, or sweat, for example. The samples may be derived from singular subjects or pools of singular samples from multiple subjects . In this example the healthy biofluid was urine.
2) the healthy urine sample was analysed to produce a healthy peptide fingerprint; We define analysed for example, as any combination of the steps of sample selection, preparation, seperation, identification, annotation, retrieval of stored data, and comparisons of data from the body of this application, the examples provided in this application, the claims of this application. We define fingerprint in this example as a identifiable singular peptide constituent of a native protein, and/or, which may be or not combined with other identifiable singular peptide constituent of a native protein, and or the sum total of all identifiable singular peptides which can be grouped together and as such that grouping becomes an entity itself. We define identifiable as the fraction, mass, and intensity of singular peptide entities.The further define mas as the unique MS mass assignment of an identifiable singular peptide, the derived mass of collected identifiable singular peptide entities, and or the derived mass of collected identifiable singular peptide entities in combination.
3) A diseased human bio fluid sample was obtained from a diseased individual; Diseased in this example is defined as an adult person aged 20-80, with or without clinical symptoms or presentation of disease, and/or with a clinical at risk

Ink. t. Patent- och reg.verket

2003 -09- 2 5

behaviour for developing disease such as smoking, drug abuse, alcoholism, overweight, with or without a diagnosed genetic disposition for disease in later onset of life. For example, patients with a clinical diagnoses of COPD, or patients at risk for developing (COPD); We define at risk for disease for example, as a current smoker of tobacco or other medicinal herb, or a person who has ever smoked, or as a person who has smoked and quit smoking, irrespective of time frame in relation to the sampling of these same patients for study. We further define at risk for disease as any person with deficiencies in the expression or the regulation of expression of alpha-1-anti-trypsin or related naturally occurring biochemical molecules, or any or any biological entity related to the expression or function of alpha-1-anti-trypsin or related naturally occurring biochemical molecules. The diseased individual in this example was a subject who showed signs of the onset or progression of Chronic obstructive Pulmonary Disease (COPD) . We further may characterize COPD patients as subjects which show elastin breakdown using histological analysis, or immunohistochemistry analysis of pulmonary tissue samples (Figure 7). We further may characterize COPD patients as subjects which show alveolitis, airway hyperinflation, or emphysema using histological analysis, or immuno-histochemistry analysis of pulmonary tissue samples. We further may characterize COPD patients as subjects which show evidence of activated macrophages within pulmonary tissue samples using histology and immunohistochemistry with antibodies specific for the detection of products of genes expressed by actoivated macrophages. In this example the patient showed histological evidence of pulmonary airwayspace enlargement, emphysema, and destruction of pulmonary elastin integrity. The example further showed evidence of activated alveolar macrophages near sites of pulmonary elastin destruction (Fig 8)

We define diseased biofluid here as any sample of human clinical material in solution form taken from patients who fulfil all or parts of the criteria of the definition of disease as above. This may include blood, serum, plasma, saliva, lavages, tears, urine, seminal fluid, joint fluid, aqueous humor, washings of cavities

or sinuses, the soluble form of tissue preparations, the soluble form of organ preparations, or sweat, for example. The samples may be derived from singular patients or pools of singular samples from multiple patients. In this example the biofluid was urine.

4) The diseased sample was analysed to produce a diseased peptide fingerprint; We define fingerprint in this example as a identifiable singular peptide constituent of a native protein, and/or, which may be or not combined with other identifiable singular peptide constituent of a native protein, and or the sum total of all identifiable singular peptides which can be grouped together and as such that grouping becomes an entity itself. We define identifiable as the fraction, mass, and intensity of singular peptide entities.The further define mas as the unique MS mass assignment of an identifiable singular peptide, the derived mass of collected identifiable singular peptide entities, and or the derived mass of collected identifiable singular peptide entities in combination.

Summary of Experimental procedure

1) a healthy human bio fluid sample was obtained;
2) the healthy urine sample was analysed to produce a healthy peptide fingerprint
3) A diseased human bio fluid sample was obtained from a diseased individual;
4) The diseased sample was analysed to produce a diseased peptide fingerprint;
5) the healthy peptide fingerprint was compared to the diseased peptide fingerprint to identify the set of peptides found only in the diseased peptide fingerprint;
6) The diseased set of peptides identified in step (5) was compared with the peptide fingerprint of the degradation products produced in step (8).

Description of the experimental procedure

The specific experimental conditions used in order to generate differentially displayed peptides in human urine samples from healthy and COPD patients was made as follows; The biofluid was sampled from patients and thereafter aliquoted and frozen at -80°C.

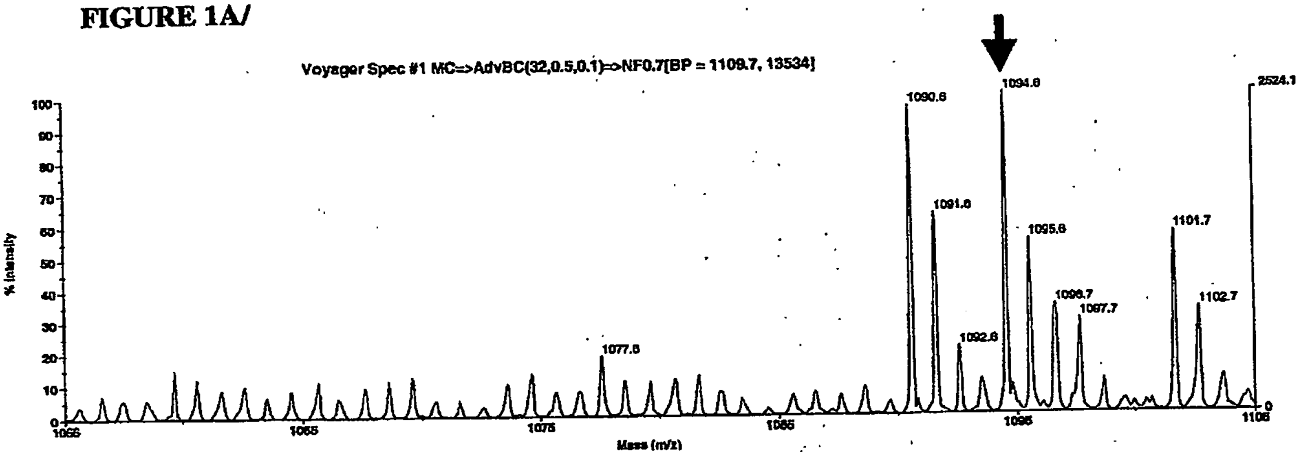
The frozen urine was thawed in room temperature, pH adjusted to 2.5 with orthophosphoric acid and processed for HPLC separations. The urine samples were introduced into a two-dimensional chromatography system where the first separation mechnism utilised was size exclusion chromatography. The cut-off of the column material was approximately 15 kDa. The fractionations resulting from the size exclusion separation step is next transferred on-line to a cation –exchange step where the peptides/proteins will be separated based upon charge. These fractions are then transferred to a reversed phase separation step where all interfering matrix components present in the sample is eliminated, this is the third dimensional separation. The third dimension fractions are the spotted down onto a MALDI target plate by a robotic feeder that adds the MALDI matrx to the peptide/årootein sample sptos. Next the MALDI sample paltae is inserted in the MALDI-TOF mass spectrometer instrument and analysed according the exact mass, quantity and isotope resolution.

Figure 1 and Figure 2 illustrate two examples where we have qualitative and quantitative differences in-between helathy and COPD patient data from resulting Elsatin peptides in the urine samples, respectively. In Figs 1A and 2A, the biomarker peptides are found present in the biofluid from COPD patients, due to active MMP-12 degradation, while Figures 1B and 2B do not hold these peptides. The absence of these peptides in the healthy subjects were not possible to quantitate (found absent) at the lowest level of quantification that MALDI instrumentation offers.

Results are shown in Figure 1, which presents the mass spectrum generated from a given liquid phase separation fraction with peptides having a specific given chemical-physical property peptide. In Figure 1A the resulting mass spectrum shows the presence of the mass peak of 1094.60 in COPD patients that corresponds to a MMP-12 enzymatic product with elastin as a substrate. Figure 1B illustrates the resulting mass spectrum from healthy volunteers where the 1094.60 mass of the peptide peak is not qualitatively distinguishable from the background signals. These data are statistically significant when compared within the diseased and healthy patient groups.

Figure 1 shows mass spectrum generated from a given liquid phase separation fraction where; A/ present in a human urine fraction from a COPD sufferer, B/ absent in a healthy subject, and C/MMP-12 + elastin, 2.5h incubation.

FIGURE 1A/



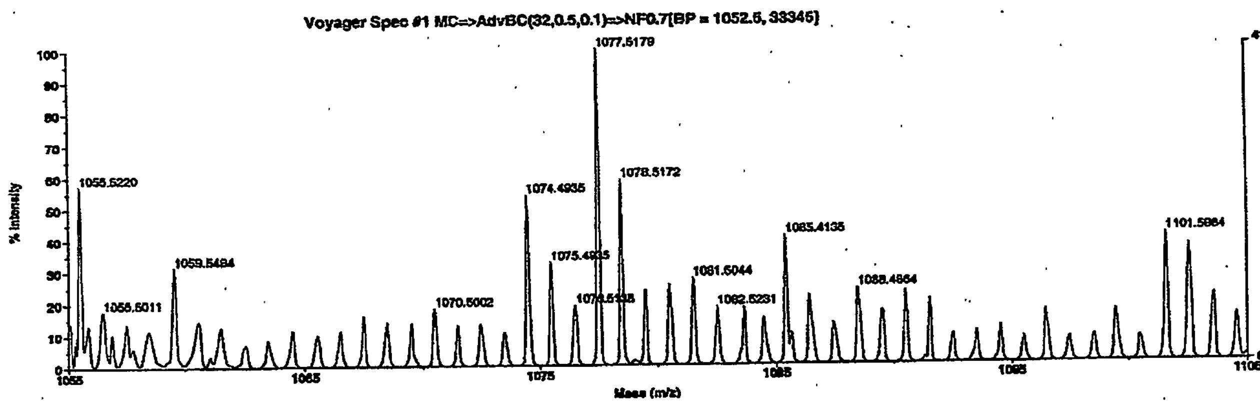
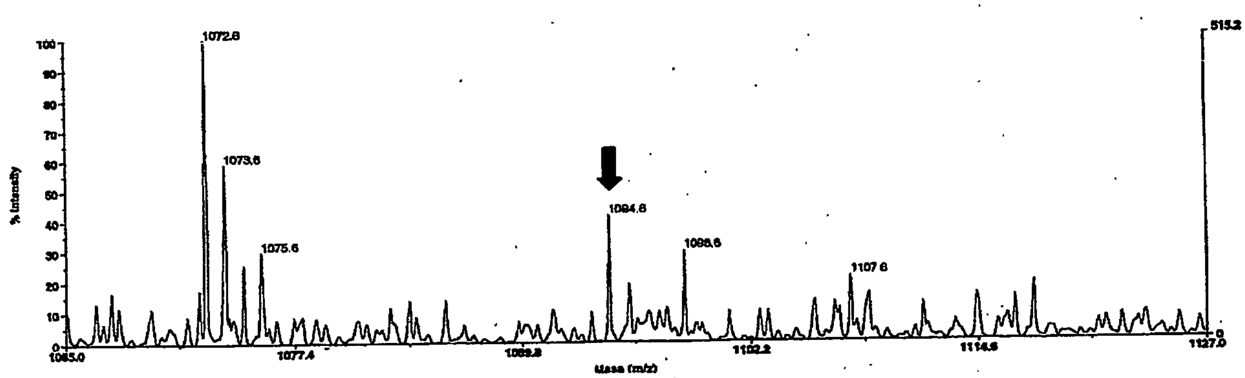


Figure 1C/



Ink. t. Patent- och reg.verket

2003 -09- 2 5

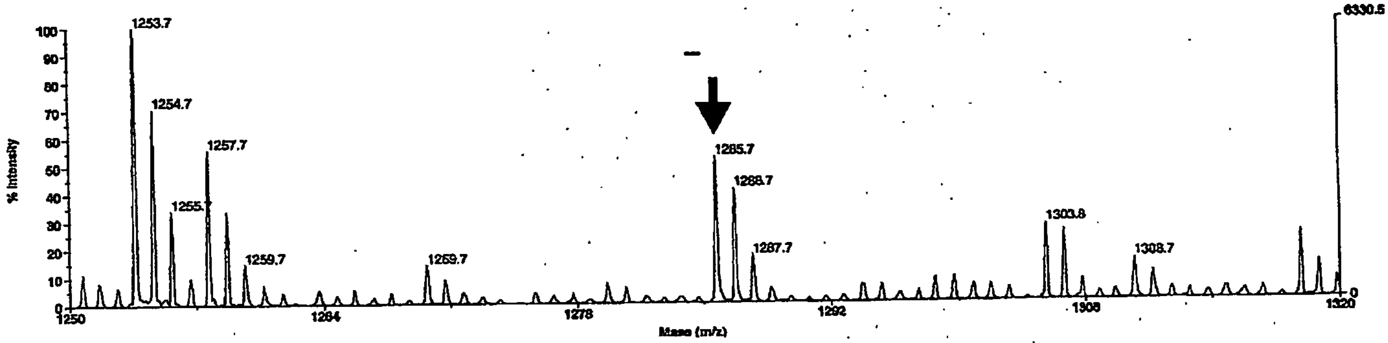
EXAMPLE 2

The following data were generated according to the method described in Example 1.

Results are shown in Figure 2 which shows that the peptide peak corresponding to a mass of 1285.7 is A/ present in a human urine fraction from a COPD sufferer, B/ absent in a healthy subject, and C/MMP-12 + elastin, 2.5h incubation

Figure 2A shows that the peptide peak corresponding to a mass of 1285.7 is present in a human urine fraction from a COPD sufferer. Figure 2B shows that this peptide peak is absent from the mass spectrum of the healthy individual. These data are statistically significant when compared within the diseased and healthy patient groups.

FIGURE 2A/



B/

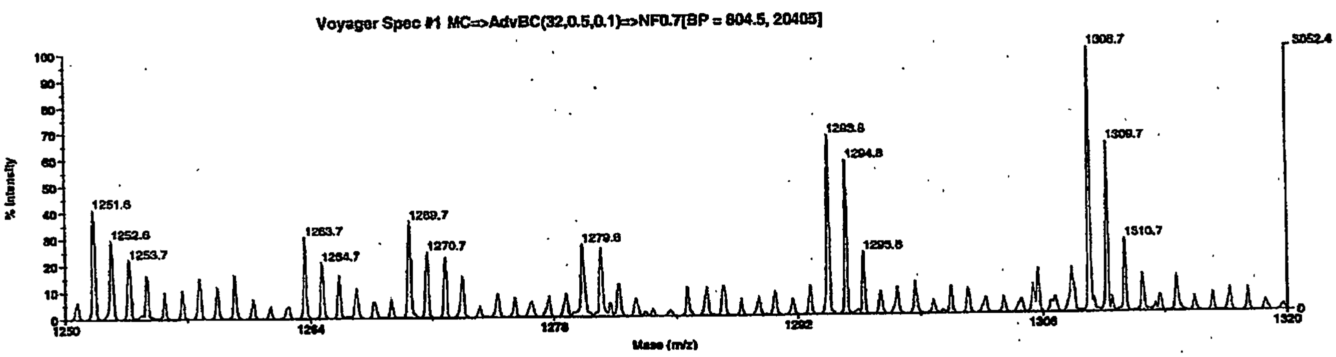
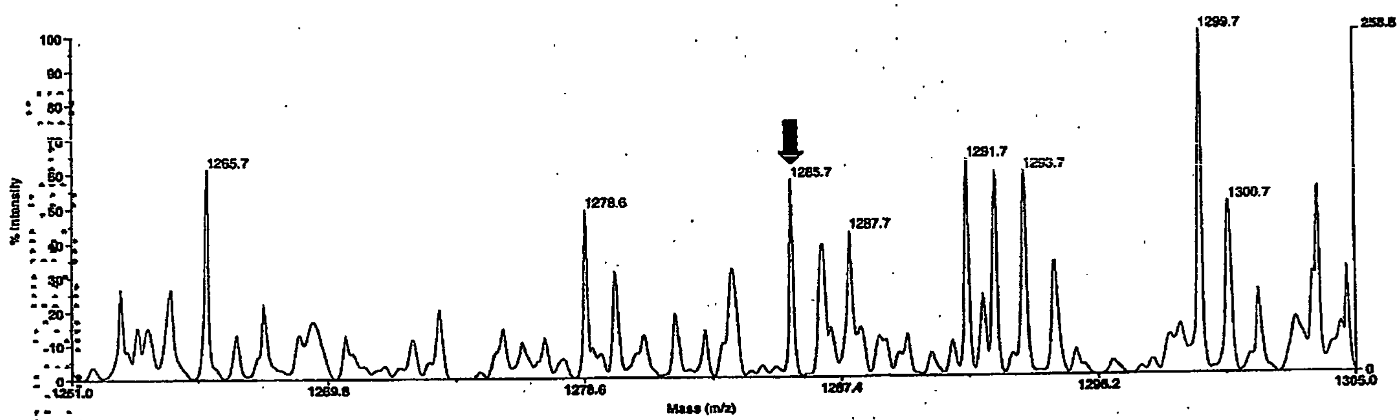


Figure 2C/



EXAMPLE 3

Direct demonstration of elastin degradation in diseased tissue

Elastin break down patterns were identified as differential expression patterns in-between healthy and diseased patients.

Elastin is a principal component of elastic fibres constituting a main part of the lung's extracellular matrix, and is known to be digested by proteolytic enzymes in emphysema. The elastin immunohistochemistry staining of COPD lungs resulted in a focal staining of the alveolar openings in the parenchyma and in a scattered distribution in the alveolar walls of both disrupted and intact septa. Elastin staining easily identified small and large vessels and capillaries throughout the lung. Parallel sheets of elastin were stained around arteries and a more spread staining of the elastic lamina was seen surrounding the vessels and bronchioles. Results showed that elastin is produced and broken down in the lungs of COPD patients during the natural disease process. Approximately two thirds of all the COPD patients and persons at risk for developing COPD showed similar profiles.

Identification and evaluation of activated inflammatory cells

At the alveolar level the macrophages are the primary defence cells with a high phagocytic capacity. The staining with antibody to CD68, a cytoplasmic granule located within the lysosome of resident and inflammatory phagocytic cells clearly identified these cells within all lung compartments. The CD68 positive cells were typical large foamy alveolar macrophages either in clusters within the alveoli or attached to the walls of the alveolar septae. Typical examples of these cells were found in bronchiolitis and alveolitis. Large clusters of cells were shown filling the airway lumen. Similar patterns were found in subjects from both patient groups indicating the severity of inflammation within the lung compartments of smokers. Macrophages were found in both bronchioles and throughout the parenchyma eighter as singular cells or in clusters. In the parenchyma the macrophages were both found in the alveolar spaces, and within or attached to the septal walls.

EXAMPLE 4

Table 1 shows the MS atomic mass unit identities of elastin peptides resulting from MMP12 digestion, and separation by column chromotagraphy. Fractions of column eluates containing separated peptides were applied to MS, identified, annotated, and then placed into a database for statistical analysis.

Table 1

TABLE 1
772.4
798.4
802.4
817.5
823.4
824.4
865.4
874.5
878.5
904.4
937.5
944.5
951.5
977.5
1027.5
1072.6
1073.6
1089.5
1094.6
1107.6

TABLE 1

1114.6
1137.6
1141.6
1142.6
1145.6
1155.6
1171.6
1185.7
1193.6
1199.6
1216.6
1232.6
1237.6
1242.7
1254.7
1255.9
1262.6
1265.7
1285.7
1287.7
1291.7
1296.7
1298.7
1299.7
1314.7
1320.8
1322.7
1331.7

TABLE 1

1347.7
1352.7
1363.7
1368.8
1369.0
1370.7
1379.6
1402.8
1424.8
1440.7
1455.8
1469.8
1477.8
1484.8
1505.8
1508.9
1519.8
1520.8
1524.8
1536.8
1542.8
1570.7
1596.8
1599.8
1613.9
1644.8
1666.9
1670.9

TABLE 1

1687.9
1696.9
1702.2
1706.8
1718.9
1758.9
1762.9
1763.9
1770.0
1832.9
1840.0
1851.0
1885.0
1920.0
1929.8
1942.0
1998.1
2168.1
2367.2
2620.3
2823.5

CLAIMS

What we claim is;

1. A method to generate a peptide fingerprint of the degradation products of a disease-associated enzyme X, wherein enzyme X is associated with disease Y, which comprises:
   (d) mixing a disease-associated enzyme X with its natural substrate *in vitro* in conditions that allow interaction between enzyme X and its substrate;
   (e) allowing the substrate to be degraded by enzyme X;
   (f) analysing the mixture to produce a peptide fingerprint of the degradation products.

2. A method as claimed in claim 1 wherein enzyme X is MMP12 and disease Y is COPD.

3. A method to determine if or confirm that an enzyme X is associated with disease Y which comprises:
   (l) obtaining a healthy biofluid sample or a healthy tissue sample;
   (m) analysing the healthy sample to produce a healthy peptide fingerprint;
   (n) obtaining a diseased biofluid sample or a diseased tissue sample, wherein the diseased sample shows signs of the onset or progression of disease Y;
   (o) analysing the diseased sample to produce a diseased peptide fingerprint;
   (p) comparing the healthy peptide fingerprint to the diseased peptide fingerprint and identifying the set of peptides found only in the diseased peptide fingerprint;
   (q) mixing enzyme X with its natural substrate *in vitro* in conditions that allow interaction between enzyme X and its substrate;
   (r) allowing the substrate to be degraded by enzyme X;
   (s) analysing the mixture to produce a peptide fingerprint of the degradation products;
   (t) comparing the diseased set of peptides identified in step (e) with the peptide fingerprint of the degradation products produced in step (h), and determining if there are statistically significant similarities between them;

Ink. t. Patent- och reg.verket

2003 -09- 25

(u) if there are statistically significant similarities between the diseased set of peptides identified in step (e) and the peptide fingerprint of the degradation products produced in step (h), concluding that enzyme X is associated with disease Y;

(v) if there are no statistically significant similarities between the diseased set of peptides identified in step (e) and the peptide fingerprint of the degradation products produced in step (h), concluding that enzyme X is not associated with disease Y.

4. A method as claimed in claim 3 wherein enzyme X is MMP12 and disease Y is COPD.

5. A method to determine the presence of a disease Y which comprises:

(i) obtaining a biofluid sample or a tissue sample;

(j) analysing the sample to obtain its peptide fingerprint;

(k) mixing enzyme X with its natural substrate *in vitro* in conditions that allow interaction between enzyme X and its substrate, wherein enzyme X is associated with disease Y;

(l) allowing the substrate to be degraded by enzyme X;

(m) analysing the mixture to produce a peptide fingerprint of the degradation products;

(n) comparing the peptide fingerprint of the sample identified in step (b) with the peptide fingerprint of the degradation products produced in step (e), and determining if there are statistically significant similarities between them;

(o) if there are statistically significant similarities between the peptide fingerprint of the sample identified in step (b) and the peptide fingerprint of the degradation products produced in step (e), concluding that disease Y is present;

(p) if there are no statistically significant similarities between the peptide fingerprint of the sample identified in step (b) and the peptide fingerprint of the degradation products produced in step (e), concluding that disease Y is absent or is being successfully treated.

6. A method as claimed in claim 5 wherein enzyme X is MMP12 and disease Y is COPD.

7. A diagnostic test kit for determining the presence of a disease Y which comprises means to compare the peptide fingerprint of a biofluid sample or the peptide fingerprint of a tissue sample with the peptide fingerprint of the degradation products in a mixture of enzyme X with its natural substrate, wherein enzyme X is associated with disease Y.

8. A method as claimed in claim 7 wherein enzyme X is MMP12, its natural substrate is elastin, and disease Y is COPD.

9. A method to analyse the effect of a drug Z on enzyme X, wherein enzyme X is associated with disease Y, which comprises:

(i) treating a human or non-human animal with the drug Z, wherein the human or non-human animal is suffering from disease Y;

(j) obtaining a biofluid sample or a tissue sample from the human or non-human animal;

(k) analysing the sample to obtain its peptide fingerprint;

(l) mixing enzyme X with its natural substrate *in vitro* in conditions that allow interaction between enzyme X and its substrate, allowing the substrate to be degraded by enzyme X;

(m) analysing the mixture to produce a peptide fingerprint of the degradation products;

(n) comparing the peptide fingerprint of the sample identified in step (c) with the peptide fingerprint of the degradation products produced in step (e), and determining if there are statistically significant similarities between them;

(o) if there are statistically significant similarities between the peptide fingerprint of the sample identified in step (c) and the peptide fingerprint of the degradation products produced in step (e), concluding that drug Z is not inhibiting enzyme X;

(p) if there are no statistically significant similarities between the peptide fingerprint of the sample identified in step (c) and the peptide fingerprint of the degradation products produced in step (e), concluding that drug Z is inhibiting enzyme X.

10. A method as claimed in claim 9 wherein enzyme X is MMP12 and disease Y is COPD.

11. A diagnostic test kit for analysing the effect of a drug Z on enzyme X which comprises means to compare the peptide fingerprint of a biofluid sample or the peptide fingerprint of a tissue sample with the peptide fingerprint of the degradation products in a mixture of enzyme X with its natural substrate, wherein the sample has been obtained from a human or non-human animal that has been or is being treated with the drug Z. Preferably enzyme X is MMP12.

ABSTRACT

Methods for producing and using protein/peptide fingerprints, allowing identification and investigation of disease-associated proteins/peptides that can be linked to specific drug targets, or to specific drug target combinations. The methods are particularly useful for studies relating to Chronic Obstructive Pulmonary Disease (COPD), especially for the enzyme MMP12.